

Tryptorubin A: A Polycyclic Peptide from a Fungus-Derived Streptomycete

Thomas P. Wyche,[†] Antonio C. Ruzzini,[†] Laura Schwab,[‡] Cameron R. Currie,[‡] and Jon Clardy^{*,†}

[†]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, United States

[‡]Department of Bacteriology, University of Wisconsin–Madison, 1550 Linden Drive, Madison, Wisconsin 53706, United States

S Supporting Information

ABSTRACT: Fungus-growing ants engage in complex symbiotic relationships with their fungal crop, specialized fungal pathogens, and bacteria that provide chemical defenses. In an effort to understand the evolutionary origins of this multilateral system, we investigated bacteria isolated from fungi. One bacterial strain (*Streptomyces* sp. CLI2509) from the bracket fungus *Hymenochaete rubiginosa*, produced an unusual peptide, tryptorubin A, which contains heteroaromatic links between side chains that give it a rigid polycyclic globular structure. The three-dimensional structure was determined by NMR and MS, including a ¹³C-¹³C COSY of isotopically enriched material, degradation, derivatives, and computer modeling. Whole genome sequencing identified a likely pair of biosynthetic genes responsible for tryptorubin A's linear hexapeptide backbone. The genome also revealed the close relationship between CLI2509 and *Streptomyces* sp. SPB78, which was previously implicated in an insect–bacterium symbiosis.

Natural product studies largely focus on their therapeutic potential, but studies motivated by their ecological roles, which place them in their evolutionary context, have become increasingly rewarding for chemistry and biology.^{1–4} For the last several years, our laboratories have collaborated to understand the molecular basis underlying the complex relations between fungus-growing ants, their fungal crop, their crop's specialized fungal pathogen, and bacterial symbionts that provide chemical defenses.^{4–7} Examples such as dentigerumycin,⁴ 9-methoxyrebeccamycin,⁵ and selvamicin⁶ illustrate some of the molecular and biological diversity discovered in this system.

The complex multilateral symbiotic system seen today undoubtedly had its evolutionary origin in a simpler system. Two types of fungi, a basidiomycete that would eventually emerge as the fungal crop and an ascomycete that would eventually emerge as the specialized pathogen, and their antagonistic interactions are a critical component of the host microbiome. A plausible origin begins with a fungus associating with a bacterium (actinomycete) to provide chemical defenses and the subsequent recruitment of fungus–bacterium pair by an ant. The ascomycete pathogen followed, and the ant–fungus–bacterium mutualism continuing to evolve and diversify for ~50 million years.

To investigate the plausibility of a fungus–bacterium origin, actinobacterial isolates (e.g., *Streptomyces* spp.) were obtained from a fungi collection, and extracts of these strains were analyzed by LCMS. One strain, *Streptomyces* sp. CLI2509 (isolated from the bracket fungus *Hymenochaete rubiginosa*) produced a compound with an *m/z* 897.3928 [M + H]⁺, a mass that did not match any compounds in the natural products database AntiBase. CLI2509 was selected for large-scale fermentation, and the molecule with *m/z* 897 was purified and structurally characterized. HRMS data provided a molecular formula of C₄₉H₅₂N₈O₉ for the compound, which we have named tryptorubin A (1; Figure 1). Initial analysis of

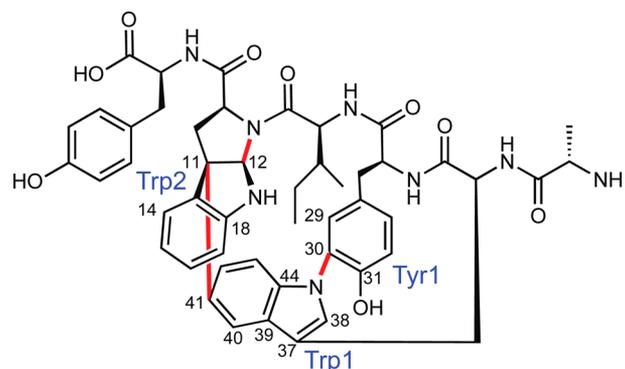


Figure 1. Structure of tryptorubin A (1).

1D and 2D NMR data indicated that the compound was a hexapeptide with characteristic ¹H and ¹³C NMR shifts for alanine, isoleucine, and tyrosine, but the remaining three amino acids could not be easily identified due to unusual NMR shifts: C-29/H-29 (δ_C 131.8, δ_H 5.76) and C-12/H-12 (δ_C 92.3, δ_H 6.78). Additionally, there appeared to be aromatic bridging between these remaining three amino acids as indicated by an HMBC correlation (Figure 2) between H-40 (δ_H 7.17) and C-11 (δ_C 60.9). There were also several heteroatoms present as indicated by atypical downfield ¹³C NMR shifts. Tryptorubin A (1) has ten exchangeable protons as shown by dissolution in CD₃OD:D₂O (1:1) and analysis by HRMS.

Because of the unusual NMR shifts and aromatic bridging between amino acids, several steps were needed to characterize the three remaining amino acids. First, small-scale (25 mL)

Received: June 14, 2017

Published: August 30, 2017

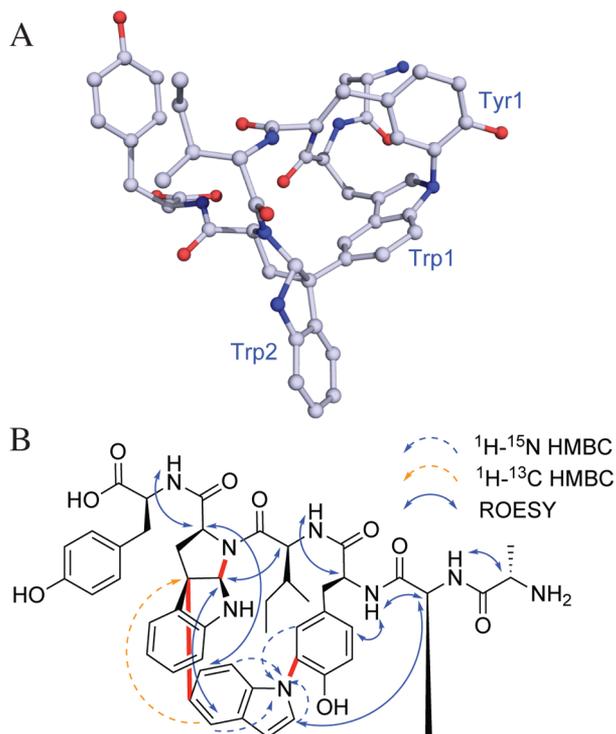


Figure 2. (A) 3D model of tryptorubin A. (B) Key 2D NMR correlations for tryptorubin A.

cultures of strain CLI2509 were grown with ^{15}N -labeled amino acids, one amino acid per culture, and the extract was analyzed by LCMS to determine which amino acids were incorporated and in what quantity. Alanine, isoleucine, and tyrosine were used as positive controls, and tryptophan, phenylglycine, and phenylalanine were also tested. LCMS analysis of these labeled cultures determined that tryptorubin A (**1**) contained alanine, isoleucine, two tryptophans, and two tyrosines (see [Supporting Information](#)). To clarify the location of the tryptophans, CLI2509 was grown with deuterium-labeled tryptophan (indole- d_3), and ^1H NMR and HRMS analysis of the purified product determined nine deuterium atoms had replaced hydrogen atoms (H-12, H14–H17, H-38, H-40, H-42, H-43; see [Supporting Information](#)). Finally, to confirm the incorporation of two tyrosines, fermentation of strain CLI2509 (500 mL) in ISP2 containing 300 mg labeled L-tyrosine (ring-3,5- d_2) led to the production of tryptorubin A (**1**) containing three deuterium atoms as evidenced by LCMS and NMR and confirming the branching at C-30. These three experiments helped determine the location of the two tryptophans, and an HMBC correlation from H-40 (δ_{H} 7.17) to C-11 (δ_{C} 60.9) suggested that the tryptophans were connected to each other by a carbon–carbon bond between C-11 and C-41. While this linkage is unusual in natural or unnatural molecules, for that matter, a similar structural motif is present in natural products such as naseseazine A,⁸ asperazine,⁹ and pestalazine A.¹⁰

After identifying the location of the two tryptophans, the location of the last amino acid with aromatic bridging, tyrosine, was determined by isotopic labeling and NMR analysis. C-30 and C-31 were not bonded to hydrogen atoms, but the connectivity of these atoms was unclear due to the ambiguous NMR shifts in the aromatic ring. Consequently, isotopic labeling (^{15}N and ^{13}C) of tryptorubin A (**1**) was used to rapidly complete the structure. Fermentation of strain CLI2509 in ISP2

medium containing ^{15}N -labeled NH_4Cl produced ^{15}N -labeled tryptorubin A (**1**). ^{15}N HMBC analysis of ^{15}N -labeled tryptorubin A (**1**) provided evidence of the linkage between tryptophan and the tyrosine-derived amino acid with HMBC correlations from H-29 (δ_{H} 5.76), H-38 (δ_{H} 6.89), H-40 (δ_{H} 7.17), and H-43 (δ_{H} 7.06) to the indole nitrogen, suggesting that the indole nitrogen was connected to C-30. The ^{13}C NMR shift at C-31 (δ_{C} 149.6), suggested that it was connected to an oxygen. In order to confirm the carbon connectivity in these two amino acids, fermentation of strain CLI2509 in ISP2 medium containing ^{13}C -labeled glucose produced ^{13}C -labeled tryptorubin A (**1**). A ^{13}C - ^{13}C COSY of ^{13}C -labeled tryptorubin A (**1**) provided the carbon–carbon connectivity for most of the structure, including the tryptophan and tyrosine.¹¹

To confirm the location of the hydroxyl groups, tryptorubin A (**1**) was acetylated (see [Supporting Information](#)). Acetylation of tryptorubin A (**1**) resulted in a product with three acetyl groups, but NMR analysis did not allow the unequivocal location of the acetyl groups to be determined. In an alternative approach, tryptorubin A (**1**) was methylated, and the major product contained the *O*-methyl at C-1, as evidenced by 1D and 2D NMR analysis. The hydroxyl at C-31 was confirmed by TOCSY correlations from 31-OH (δ_{H} 7.28) to H-33. The hydroxyl at C-7 was confirmed by a comparison of ^{13}C and ^1H NMR shifts to literature values. After the amino acid identities and aromatic bridging were determined, analysis of ROESY, COSY, and HMBC NMR data allowed for determination of the sequence of tryptorubin A (**1**) ([Figure 2](#)).

A combination of advanced Marfey's method, genome sequencing, NOE correlations, and molecular modeling assigned the absolute stereostructure of tryptorubin A (**1**). Acid hydrolysis of tryptorubin A (**1**) and subsequent derivatization with L-FDLA and DL-FDLA and LCMS^{12,13} analysis provided the configuration of L-Ala, L-Ile, and L-Tyr (see [Supporting Information](#)). Whole genome sequencing of strain CLI2509 and identification of the gene cluster responsible for producing tryptorubin A (**1**) indicated that all of the 6 amino acids had an *L*-configuration. The configurations at C-11 and C-12 remained unassigned, leaving four possible stereoisomers. The unusual branching between aromatic regions of the remaining amino acids prevented determination of the configuration of these two stereocenters using Marfey's method. Instead, several key ROESY correlations ([Figure 2B](#)) existed in tryptorubin A (**1**) that helped determine the stereochemistry: between H-9 (δ_{H} 4.52) and H-42 (δ_{H} 7.44), as well as H-12 (δ_{H} 6.78) to H-40 (δ_{H} 7.17), suggesting that H-9, H-12, and H-42 were on the same side of the tryptophan. The four possible stereoisomers were modeled using Schrödinger and Gaussian09 software and only one of the four stereoisomers (11*S*, 12*R*) fit with the experimental ROESY correlations.¹⁴ Consequently, the absolute stereochemistry was determined to be (2*S*, 9*S*, 11*S*, 12*R*, 20*S*, 21*S*, 26*S*, 35*S*).

Tryptorubin A (**1**) has several unusual features, some previously reported and some unreported. The linkage between two tryptophans, the C-11 to C-41 in tryptorubin A (**1**), is rare, but similar links are found in naseseazine A, asperazine, and pestalazine A. However, in tryptorubin A (**1**) the linkage is *para* to the indole nitrogen; in naseseazine and pestalazine A, the linkage is *meta* to the nitrogen; and in asperazine, the linkage is *ortho* to the nitrogen. Another unusual structural feature in tryptorubin A (**1**) is the linkage between tyrosine's aromatic ring (C-30) to the indole nitrogen of tryptophan, a feature not reported in any other natural product. An alkaloid produced by

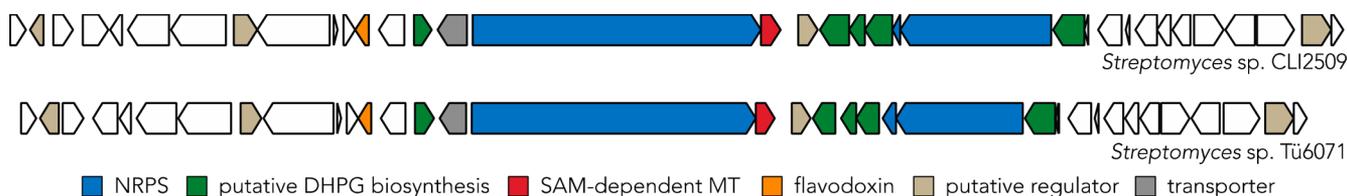


Figure 3. antiSMASH-predicted hybrid type III PKS/NRPS biosynthetic gene clusters in *Streptomyces* sp. CLI2509 and Tü6071. The NRPS genes (blue) are predicted to encode for a hydrophobic hexapeptide, and we have tentatively assigned these genes to tryptorubin A peptide synthesis (for bioinformatic annotation, see the [Supporting Information](#)).

*Penicillium citreo-viride*¹⁵ has a related linkage (see [Supporting Information](#)) that is likely produced in a different fashion. A linkage from the indole nitrogen in tryptophan to other amino acids is not unprecedented; pestalazine B, aspergilazine A,¹⁶ and kapakahine B,¹⁷ for example, are bonded between the indole nitrogen to another tryptophan. Tryptorubin A's unusual structural features and the lack of annotated biosynthetic pathways for any structural relatives prompted an investigation of its genetic basis. We began our genetic analysis of tryptorubin A (**1**) by sequencing and analyzing the *Streptomyces* sp. CLI2509 genome. Assembly of PacBio sequence reads resulted in two linear replicons: a 7.09 Mb chromosome and a 147 kb plasmid (Genbank accession no. CP021118 and CP021119). The chromosome encodes for 18 antiSMASH-predicted biosynthetic gene clusters (BGCs),¹⁸ many of which we noticed were also predicted in two other bacteria: *Streptomyces* sp. strain Tü6071 (Genbank accession no. CM001165.1), isolated from the soil along the Cape Coast in Ghana,¹⁹ and *Streptomyces* sp. SPB78,^{20,21} an antifungal producer isolated from the Southern Pine Beetle *Dendroctonus frontalis*. A comparison of the three bacterial chromosomes using *in silico* genome-to-genome distance calculations²² revealed that CLI2509 was remarkably similar to both Tü6071 and SPB78, having predicted DNA–DNA hybridization values of 91.3% and 77–82.6%, respectively. In light of the genome sequence similarity, and our previous work on *Streptomyces* sp. SPB78 while studying a fungus-growing beetle system, we investigated this bacterium's ability to produce tryptorubin A (**1**). The cultivation of SPB78 on ISP2 confirmed its ability to produce tryptorubin A (**1**) as determined by HRMS.

Among the predicted BGCs encoded on the CLI2509 chromosome, none can be confidently predicted to encode for the biosynthesis of tryptorubin A (**1**). Only a single locus, which encodes for two NRPSs that fulfill the criteria for hexapeptide formation (Figure 3) exists, and we speculate it is involved in tryptorubin A production. The rest of this BGC, however, presents several curiosities, including the sequences for enzymes that are clearly not involved in the biosynthesis of tryptorubin A (**1**). For example, tryptorubin A (**1**) is built from only proteinogenic amino acids yet the biosynthetic locus encodes for a set of enzymes similar to those required for dihydroxyphenylglycine (DHPG) biosynthesis. In addition, two SAM-dependent enzymes are present: the first is a predicted *N*-methyltransferase NRPS module and the second is a stand-alone methyltransferase. Finally, among the remaining genes in the neighborhood, one encodes for a flavodoxin, a common radical SAM redox partner, but its partner is unknown.

The missing genes, encoded elsewhere on the chromosome, cannot be identified using a bioinformatics approach. These encode for tryptorubin A's unusual side chain cyclization. Two C–N bonds and one C–C bond must be installed to reach the

final product. There are presumably at least two enzymes that accomplish these reactions: one that forms a C–N bond between the indole *N* ϵ of Trp1 and a Tyr1 *C* ϵ 2 (C-30), and a second that catalyzes C–C bond formation between the indole rings at Trp1 *C* ζ 3 (C-41) and Trp2 *C* γ (C-11). The latter is likely concomitant with C–N bond and ring formation between *C* δ 1 (C-12) and the Trp2 amide. These two putative enzymes most likely carry out one electron oxidations that would allow all three bonds to be formed in a largely precedented fashion,²³ but the genes responsible for these reactions are not obvious and merit further investigation.

In summary, a study on the origin of a complex symbiosis led to a *Streptomyces*, isolated from the fungus *Hymenochaete rubiginosa* that produced a new peptide, tryptorubin A (**1**). Its fascinating structural features provide a chemical rationale for further studies on this and related fungus-hosted bacteria, and have generated an interesting biosynthetic puzzle: the enzymes that are responsible for the remarkable cyclization reactions. The results also reinforce the need for continued molecular analyses (even in well-studied taxa like *Streptomyces*) as our bioinformatic predictive abilities continue to lag behind the diversity of chemistry that is genetically encoded by bacteria. Finally, the similarity of CLI2509 to SPB78 from the Southern Pine beetle system²⁰ hints at the recruitment of another fungus–bacterium as the origin of another fungus-growing insect system. The identification of similar genomes and the same rare molecule in two different fungus-growing systems could, of course, be a coincidence, but it provides motivation for further studies.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b06176.

1D and 2D NMR spectra, molecular modeling data, genome summary, cluster annotation, and additional experimental information (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*jon_clardy@hms.harvard.edu

ORCID

Thomas P. Wyche: 0000-0003-4703-741X

Jon Clardy: 0000-0003-0213-8356

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by funding from the NIH R01 GM086258 (J.C.) and NIH U19 AI09673 (J.C., C.R.C.). A.C.R. was supported by a Harvard Medical School–Merck

Fellowship and a CIHR Fellowship. We thank the ICCB-Longwood Analytical Chemistry Core Facility at Harvard Medical School for the facilities to acquire LCMS data, as well as the East Quad NMR Facility at Harvard Medical School for the facilities to acquire NMR data. We also thank Arvin Moser and Maria Varlan (ACD/Labs) for assistance with NMR software. We thank the Duke GCB Genome Sequencing Shared Resource, which provided PacBio sequencing service, the Harvard Medical School Information Technology Department for access to the Orchestra High Performance Computing Cluster, and the Harvard University Information Technology Department for access to the Odyssey High Performance Computing Cluster. We thank Emily Mevers for measurement of optical rotation.

REFERENCES

- (1) Yim, G.; Wang, H. H.; Davies, J. *Philos. Trans. R. Soc., B* **2007**, *362*, 1195–1200.
- (2) Flórez, L. V.; Scherlach, K.; Gaube, P.; Ross, C.; Sitte, E.; Hermes, C.; Rodrigues, A.; Hertweck, C.; Kaltenpoth, M. *Nat. Commun.* **2017**, *8*, 15172.
- (3) van der Meij, A.; Worsley, S. F.; Hutchings, M. I.; van Wezel, G. P. *FEMS Microbiol. Rev.* **2017**, *41*, 392–416.
- (4) Oh, D. C.; Poulsen, M.; Currie, C. R.; Clardy, J. *Nat. Chem. Biol.* **2009**, *5*, 391–393.
- (5) Van Arnam, E. B.; Ruzzini, A. C.; Sit, C. S.; Currie, C. R.; Clardy, J. *J. Am. Chem. Soc.* **2015**, *137*, 14272–4.
- (6) Van Arnam, E. B.; Ruzzini, A. C.; Sit, C. S.; Horn, H.; Pinto-Tomás, A. A.; Currie, C. R.; Clardy, J. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12940–12945.
- (7) Carr, G.; Derbyshire, E. R.; Caldera, E.; Currie, C. R.; Clardy, J. *J. Nat. Prod.* **2012**, *75*, 1806–9.
- (8) Raju, R.; Piggott, A. M.; Conte, M.; Aalbersberg, K. F.; Feussner, K.; Capon, R. J. *Org. Lett.* **2009**, *11*, 3862–3865.
- (9) Varoglu, M.; Corbett, T. H.; Valeriote, F. A.; Crews, P. *J. Org. Chem.* **1997**, *62*, 7078–7079.
- (10) Ding, G.; Jiang, L.; Guo, L.; Chen, X.; Zhang, H.; Che, Y. *J. Nat. Prod.* **2008**, *71*, 1861–1865.
- (11) Reibarkh, M.; Wyche, T. P.; Sauri, J.; Bugni, T. S.; Martin, G. E.; Williamson, R. T. *Magn. Reson. Chem.* **2015**, *53*, 996–1002.
- (12) Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591–596.
- (13) Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. *Anal. Chem.* **1997**, *69*, 5146–5151.
- (14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (15) Vining, L. C.; McInnes, A. G.; Smith, D. G.; Wright, J. L. C.; Taber, W. A. *FEMS Symp.* **1982**, *13*, 243–251.
- (16) Cai, S.; Kong, X.; Wang, W.; Zhou, H.; Zhu, T.; Li, D.; Gu, Q. *Tetrahedron Lett.* **2012**, *53*, 2615–2617.
- (17) Nakao, Y.; Yeung, B. K. S.; Yoshida, W. Y.; Scheuer, P. J.; Kelly-Borges, M. *J. Am. Chem. Soc.* **1995**, *117*, 8271–8272.
- (18) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. *Nucleic Acids Res.* **2015**, *43*, W237–W243.
- (19) Gebhardt, K.; Meyer, S. W.; Schinko, J.; Bringmann, G.; Zeeck, A.; Fiedler, H.-P. *J. Antibiot.* **2011**, *64*, 229–232.
- (20) Scott, J. J.; Oh, D.-C.; Yuceer, M. C.; Klepzig, K. D.; Clardy, J.; Currie, C. R. *Science* **2008**, *322*, 63.
- (21) Blodgett, J. A.; Oh, D.-C.; Cao, S.; Currie, C. R.; Kolter, R.; Clardy, J. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 11692–11697.
- (22) Auch, A. F.; Von Jan, M.; Klenk, H.-P.; Göker, M. *Stand. Genomic Sci.* **2010**, *2*, 117–134.
- (23) Tang, M.-C.; Zou, Y.; Watanabe, K.; Walsh, C. T.; Tang, Y. *Chem. Rev.* **2017**, *117*, 5226–5333.